

Linear Algebra

haruki

September 27, 2024

Contents

1	Linear Equations	1
1.1	What is Fields?What is Vector Spaces?	1
1.2	Systems of Linear Equations	3
1.3	Matrices and Elementary Row Operations	3
2	Vector Spaces	8
2.1	Subspaces,Basis and Dimension	8
2.2	Coordinate and Rank	10
3	Linear Transformations	13
3.1	Linear Transformations and Matrices	13

1 Linear Equations

1.1 What is Fields?What is Vector Spaces?

In short,a Field F should satisfy these properties:

(A) Axioms for addition

(A1) if x,y in F ,then $x + y$ in F

(A2) Addition is commutative:

$$x + y = y + x$$

for all x and y in F

(A3) Addition is associative:

$$x + (y + z) = (x + y) + z$$

for all x,y and z in F

(A4) There is an element 0 in F such that $x + 0 = x$,for all x in F

(A5) To every x in F , there exists an element $-x$ in F such that

$$(-x) + x = 0$$

(M) Axioms for multiplication

(M1) if x, y in F , then xy in F

(M2) Multiplication is commutative:

$$xy = yx$$

for all x and y in F

(M3) Multiplication is associative:

$$x(yz) = (xy)z$$

for all x, y and z in F

(M4) There is an element $1 \neq 0$ in F such that $1x = x$, for all x in F

(M5) To every $x \neq 0$ in F , there exists an element $\frac{1}{x}$ in F such that

$$\left(\frac{1}{x}\right)x = 1$$

(D) The distributive law

$$x(y + z) = xy + xz$$

for all x, y and z in F

Some familiar propositions like cancellation law and the uniqueness of 0 and 1 is omitted here. See the comprehensive proofs in Baby Rudin.

In the following context, we assume the field F as a subfield of the complex numbers. Otherwise, it could be possible that $1 + 1 + \cdots = 0$! You must not want to see that, which arises from the nonzero characteristic of the field. So we assume the field F we discuss is a field of characteristic zero.

Our preparation work for vector spaces has done. Let's define it!

Definition 1.1. A vector space consists of the following:

1. a field F for scalars
2. a set V of objects, called vectors
3. an operation called vector addition, which satisfies the axioms for additions (A)
4. an operation called scalar multiplication, which satisfies:
 - (a) $c\alpha$ is in V for every α in V
 - (b) $1\alpha = \alpha$

- (c) $(c_1 c_2)\alpha = c_1(c_2)\alpha$
 (d) $c(\alpha + \beta) = c\alpha + c\beta$
 (e) $(c_1 + c_2)\alpha = c_1\alpha + c_2\alpha$

The scalar multiplication has familiar properties just like the ordinary multiplication. We will dig into this concept more deeply in the following sections.

1.2 Systems of Linear Equations

Of Course, the numbers we involved are the elements of a field F .

A system of linear equations is defined as:

$$\begin{array}{ccccccc} A_{11}x_1 + A_{12}x_2 + \dots + A_{1n}x_n & = & y_1 \\ A_{21}x_1 + A_{22}x_2 + \dots + A_{2n}x_n & = & y_2 \\ \vdots & & \vdots & = & \vdots \\ A_{m1}x_1 + A_{m2}x_2 + \dots + A_{mn}x_n & = & y_m \end{array}$$

Obviously, there's no doubt that all the linear combinations of the equations are still linear equations, and every solution of the previous system of linear equations is also a solution of the new one.

However, we want to transform the original system of linear equations to the simplest form, which requires that the solutions should be the same. Therefore, if the original equations are also the linear combinations of the new one, the solutions must be the same, and we've reached our purpose. The two systems of linear equations are called **equivalent**, if each equation in each system is a linear combination of the equations in the other system. And we have the following apparent theorem.

Theorem 1.1. *Equivalent systems of linear equations have exactly the same solutions.*

Then all we have to do is to find the suitable way of linear combinations.

1.3 Matrices and Elementary Row Operations

To simplify our notations, we introduce a concept:

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix}$$

We called A a matrix. Sometimes we use the notation $\{A_{ij}\}$ to denote it.

By now the matrix is only a pile of numbers, so we endow it some operations:

1. matrix addition: $(A + B)_{ij} = A_{ij} + B_{ij}$
2. scalar multiplication: $(cA)_{ij} = cA_{ij}$
3. matrix multiplication: $(AB)_{ij} = \sum_{r=1}^p A_{ir}B_{rj}$, where AB is a $m \times n$ matrix, and A, B are $m \times p, p \times n$ matrices respectively.

The $n \times n$ identity matrix is denoted by I_n with entries I_{ij} that:

$$I_{ij} = \begin{cases} 1 & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}$$

which has a significant property: $IA = AI = A$.

The first and second operations make the matrices space become a vector space, and the third operation make the matrices space become an algebra, which will simplify our notations greatly.

An important property of matrix multiplication should be pointed out:

Proposition 1.1. *The matrix multiplication is associative.*

Proof. The proof is direct. If A, B, C are $m \times p, p \times q, q \times n$ matrices, then

$$\begin{aligned} ((AB)C)_{ij} &= \sum_{t=1}^q \left(\sum_{r=1}^p A_{ir}B_{rt} \right) C_{tj} = \sum_{t=1}^q \sum_{r=1}^p A_{ir}B_{rt}C_{tj} \\ &= \sum_{r=1}^p A_{ir} \left(\sum_{t=1}^q B_{rt}C_{tj} \right) = (A(BC))_{ij} \end{aligned}$$

□

But it isn't commutative and you can prove it directly as well.

We have possessed the matrix, so let's come back to the system of linear equations. You could find an amazing thing we expect:

$$AX = Y$$

and where

$$A = \begin{bmatrix} A_{11} & A_{12} & \dots & A_{1n} \\ A_{21} & A_{22} & \dots & A_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ A_{n1} & A_{n2} & \dots & A_{nn} \end{bmatrix}$$

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}, Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}$$

This is just the same thing we discussed above! So the objects concerned will be matrices from now on.

Don't forget the target we said at the end of the subsection 1.2! I will show you that the way we're looking for can be presented as a convenient matrix.

We define three elementary row operations of matrix as following:

1. multiplication of one row of A by a nonzero scalar c
2. replacement of the r th row of A by row r plus c times row s , c is any scalar and $r \neq s$
3. interchange of two rows of A

The elementary matrices come after:

Definition 1.2. An elementary matrix E_n is a matrix which is the consequence of an elementary row operation on the identity matrix I_n .

A simple but important conclusion reached now:

Proposition 1.2. The consequence of an elementary row operation on $m \times n$ matrix A is just $e_m A$.

Whose proof is direct and I'm lazy to type.

By the definition of elementary row operations, it's apparent that each operation has a **unique** inverse operation to cancel the effect it cause. For a further convenience, we introduce the inverse of a matrix:

Definition 1.3. Let A be a $n \times n$ matrix. If $BA = I$, we call B a **left inverse** of A . If $AC = I$, we call C a **right inverse** of A . If $BA = AB = I$, we call B a **two-side inverse** of A . You may find that:

$$B = BI = BAC = IC = C$$

So we use A^{-1} to denote the **inverse** of A , and A is called invertible.

It's simple to prove:

Theorem 1.2. 1. If A is invertible, then $(A^{-1})^{-1} = A$

2. If A, B are invertible, then AB is invertible and the inverse is $B^{-1}A^{-1}$

Then we can formalize the statement above:

Theorem 1.3. The elementary matrix is invertible, and its inverse is just the matrix which corresponds to the inverse operation.

Let's come back to the topic again. Why we choose the elementary row operations? Because we have the following theorem:

Theorem 1.4. Let A^* be (A, Y) for a system of linear equations which we call the **augmented matrix** of the system. B^*, A^* are called **row-equivalent** if they are connected by a series of elementary row operations, then the two systems are equivalent as a result.

We can write A^*, B^* in an equation:

$$B^* = E_1 E_2 \dots E_n A^{*1}$$

and an inverse one:

$$A^* = E_n^{-1} E_{n-1}^{-1} \dots E_1^{-1} B^*$$

A more convenient concept comes now:

Definition 1.4. An $m \times n$ echelon matrix R is called row-reduced if:

- (a) the first nonzero entry in each nonzero row of R is equal to 1
- (b) each column of R which contains the leading nonzero entry of some row has all its other entries 0.
- (c) if rows $1, 2, \dots, r$ are the nonzero rows of R , and if the leading nonzero entry of row i occurs in column $k_i, i = 1, 2, \dots, r$, then $k_1 < k_2 < \dots < k_r$

For example, I_n is just a row-reduced echelon matrix.

We have a theorem below, whose proof is apparent:

Theorem 1.5. Every $m \times n$ matrix over the field F is row-equivalent to a row-reduced echelon matrix.

And we have done the most of all things of systems of linear equations. You may find that if we use the elementary row operations on the augmented matrix A^* to make it become row-reduced echelon matrix R , then we get an explicit solution, because from the definition, we have x_{k_1}, x_{k_2}, \dots ² as linear combinations of the rest x_m , which by arbitrary values of x_m becomes a solution.

Of course we forgot the situation of no solutions, which can be simply demonstrated by the statement:

Proposition 1.3. Let R be the row-reduced echelon augmented matrix of a system of linear equations, then:

- If there's a first nonzero entry in the last column, there's no solutions
- If not the above condition, the solution is unique if except the last column R is I_n , otherwise there are infinite number of solutions.

If you are smart enough, you would find that the uniqueness of solutions means the coefficient matrix A is an invertible matrix, because:

$$E_1 E_2 \dots E_n A = I_n$$

implies the inverse of A is:

$$A^{-1} = E_n^{-1} E_{n-1}^{-1} \dots E_1^{-1} I_n$$

¹the foot scripts here don't mean the dimension

²the foot scripts label the first nonzero entry of each nonzero row

and the procedure can be taken by hand if we write them together (A, I_n) , and use the elementary row operations till A becomes I_n , then the right is just A^{-1} .

The last thing about systems of linear equations you need to know is that the solutions of a system of linear equations is a vector space. Let the rest x_m be a series of values:

$$(1, 0, \dots, 0), (0, 1, \dots, 0), \dots (0, 0, \dots, 1)$$

and a series of solutions come after. You may notice that each solution of the system is a linear combination of the $n - r$ solutions, which implies the solutions of the system is a vector space that has a dimension of $n - r$, which we will detail in the next section.

We finish this section by some properties of matrices:

Theorem 1.6. 1. Write B as $(\beta_1, \beta_2, \dots, \beta_n)$, then $AB = (A\beta_1, A\beta_2, \dots, A\beta_n)$

2. A is called upper triangle matrix if it only have nonzero entries above the diagonal (contain the diagonal), and the product of two upper triangle matrices is still a upper triangle matrix

3. It will be useful if you notice that AE_n corresponds to an elementary column operation, which is just a copy version of the row one.



2 Vector Spaces

2.1 Subspaces, Basis and Dimension

For a further discussion about linear algebra, we will introduce some useful concepts:

Definition 2.1. Let V be a vector space over the field F . W is a subset of V . W is a subspace of V if W is a vector space with vector addition and scalar multiplication on V .

A simple method to check W is make an arbitrary linear combination by two arbitrary vectors in W , which means:

Theorem 2.1. If α and β are vectors in W , then W is a subspace if and only if $c\alpha + \beta$ in W for each pair of α, β and each scalar c .

We will introduce the concept “span” in a new way:

Theorem 2.2. The intersection of any collections of subspaces is also a subspaces.

The proof is easy, so let's skip it.

Definition 2.2. Let S be a set of vectors of V . The **subspace spanned** by S is the intersection W of all subspaces that contains S . If S is a finite set, then we call W the **subspace spanned by the vectors** $\{\alpha_1, \alpha_2, \dots, \alpha_n\}$.

A more usual form comes from this:

Theorem 2.3. The subspace spanned by a non-empty subset S of a vector space V is the set of all linear combinations of vectors in S .

There's an operation of addition defined on the subsets:

Definition 2.3. If S_1, S_2, \dots, S_k is subsets of a vector space V , the set of all sums

$$\alpha_1, \alpha_2, \dots, \alpha_k$$

of vectors α_i in S_i is called the **sum** of the subsets S_1, S_2, \dots, S_k and is denoted by

$$S_1 + S_2 + \dots + S_k$$

If S_1, S_2, \dots, S_k are subspaces, then the sum of them is also a subspace.

A further exploration need the concept “linear dependent”:

Definition 2.4. A subset S of a vector space V is called **linearly dependent** if there exists distinct vectors $\alpha_1, \alpha_2, \dots, \alpha_n$ in S and scalars c_1, c_2, \dots, c_n in F , not all of which are 0, such that

$$c_1\alpha_1 + c_2\alpha_2 + \dots + c_n\alpha_n = 0$$

A set which is not linear dependent is called **linearly independent**. If S is only a finite set of vectors, we call these vectors **dependent** or **independent** rather than call S .

An important remark is that if S is an infinite set, the concept above didn't mean **any** infinite sum of vectors. We will see this in the following concept.

Defination 2.5. A **basis** for V is a linearly independent set of vectors in V , which span the V . The **dimension** of V is the number of elements in the basis. It's easy to see that the linear combination of each vector is unique

You may notice a theorem is a must to make the definition legitimate.

Lemma 2.1. Let $\alpha_1, \alpha_2, \dots, \alpha_k$ be independent vectors. A vector β is a linear combination of $\alpha_1, \alpha_2, \dots, \alpha_k$ if and only if $\beta, \alpha_1, \alpha_2, \dots, \alpha_k$ are dependent vectors.

Lemma 2.2. If a nonzero vector β is a linear combination of independent vectors $\alpha_1, \alpha_2, \dots, \alpha_k$, there exists i such that $\beta, \alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_k$ are independent vectors.

Proof. Because β is nonzero, there exists i such that the coefficient of α_i is nonzero, which means α_i is a linear combination of $\beta, \alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_k$. If $\beta, \alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_k$ are dependent, the coefficient of β should be nonzero otherwise $\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_k$ are dependent, which means α_i is a linear combination of $\alpha_1, \alpha_2, \dots, \alpha_{i-1}, \alpha_{i+1}, \dots, \alpha_k$, and by the Lemma 2.1, they are dependent.

Therefore the hypothesis is incorrect. \square

Theorem 2.4. If $\beta_1, \beta_2, \dots, \beta_m$ are independent vectors so as $\alpha_1, \alpha_2, \dots, \alpha_n$, and each of them is a linear combination of $\alpha_1, \alpha_2, \dots, \alpha_n$, then they can replace some of the α to form a new independent vector set.

Proof. We notice that repeat the procedure in Lemma 2.2 we can change the original set by the new set, and by the independence of β we assure the nonzero coefficients happen on α , then the proof is over. \square

From this theorem, We can prove lots of corollaries.

Corollary 2.1. If every element of a finite set of independent vectors is a linear combination of the other, then the number of elements of the former m is no more than the one of the latter n .

Corollary 2.2. The number of every finite basis is equal.

Corollary 2.3. If there exists an infinite basis, then there's no finite basis.

Hence definition of dimension is legitimate now.

A simple example of basis is below. Let V be F^{n^3} , then

$$e_i = (0, 0, \dots, 1, \dots, 0) \quad \text{The } i\text{th value is } 1$$

is a basis. It's natural to popularize this to infinite-dimensional space F^∞ , but it's wrong. Do you remember the remark above? Yes, the “**span**” should be “**finite**”

³This means n-triples on field F , whose operations defined as usual.

as the “**finite**” requirement appears in the definition of linear combination! So it’s impossible to express $(1, 1, \dots)$ by the method above. To “span” this vector space, we need **orthogonal basis** rather than usual basis, which appears in functional analysis.

How can we form a basis? A useful way to do that is extend a basis of a subspace W . The feasibility of this method is assured by Theorem 2.4, which says every basis of V can be replaced partly by a basis of W . There’s a further theorem:

Theorem 2.5. *Let W_1, W_2 be finite-dimensional subspaces of V , then*

$$\dim W_1 + \dim W_2 = \dim(W_1 \cap W_2) + \dim(W_1 + W_2)$$

Proof. We can extend the basis of $W_1 \cap W_2$ in W_1, W_2 respectively, which forms a basis of W_1, W_2 and can be proved easily. \square

Possessing these theorems, we can learn the topics about “coordinate” and “rank”, which will be explained in the next subsection.

2.2 Coordinate and Rank

For convenience, the vector space we discuss below is finite-dimensional. To describe the “coordinate”, we need a basis ordered. Having an ordered basis we can define “coordinate” as following:

Definition 2.6. *A vector α in V can be uniquely presented as a linear combination of an ordered basis $\{\alpha_i\}$, and the coefficient x_i of each α_i is called the *ith coordinate* of α related to the ordered basis $\{\alpha_i\}$. We can write the coordinates in a matrix called **coordinate matrix***

$$X = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix}$$

To do something about coordinate further, we need a theorem about matrix. You may notice we develop the properties of matrix at the time we need it.

Theorem 2.6. *Write the square matrix A as column vectors $(\alpha_1, \alpha_2, \dots, \alpha_n)$. A is invertible if and only if $\alpha_1, \alpha_2, \dots, \alpha_n$ are linearly independent.*

Proof. The proof is easy if you notice the fact that a linear combination of $\alpha_1, \alpha_2, \dots, \alpha_n$ is just a product of two matrices corresponded to a system of linear equations:

$$c_1\alpha_1 + c_2\alpha_2 + \dots + c_n\alpha_n = A \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}$$

and immediately we finish the proof. \square

Let's come back to the topic. One thing you must notice is the term "relate", which tells us the arbitrariness of the coordinate because of the arbitrary choice of basis. Therefore a natural question is how to change the coordinate when the basis changed.

Theorem 2.7. Let $\{\alpha_i\}, \{\alpha'_i\}$ be two bases. P is a matrix whose entries is

$$\alpha_i = \sum_{j=1}^n P_{ji} \alpha'_j$$

Then the two coordinates are related by

$$X' = PX$$

or

$$X = P^{-1}X'$$

The invertibility of P is guaranteed by Theorem 2.6.

Inversely, an invertible matrix P relate $\{\alpha_i\}$ to a unique basis $\{\alpha'_i\}$ that

$$X' = PX$$

and

$$X = P^{-1}X'$$

.

The proof is direct.

The rest properties of coordinate are left to the next section. Therefore let's discuss the term "rank".

Definition 2.7. The **row rank** of a matrix A is the dimension of the vector space⁴ spanned by the row vectors of A . The **column rank** is defined similarly.

Theorem 2.8. Row-equivalent matrices have the same row space.

Proof. Use Theorem 2.4. □

The similar conclusion about column-equivalent is established as well.

Theorem 2.9. The nonzero rows of row-reduced echelon matrix R is a basis of the row space. So as column-reduced echelon matrix.

If a matrix R is both a row-reduced echelon matrix and a column-reduced echelon matrix, it's direct to draw the conclusion:

Theorem 2.10. The row rank is equal to the column rank. Hence we call them the **rank** of matrix A .

We can also call the finite dimension of a space spanned by a vector set the **rank** of the vector set.

Why we introduce the rank of a matrix? You will find the answer in the next section.

⁴Obviously a subspace of F^n



3 Linear Transformations

3.1 Linear Transformations and Matrices